REMARKS

The present application was filed on June 23, 2003 with claims 1-22. Claims 1-22 remain pending, and claims 1, 10, 19, 21 and 22 are the pending independent claims.

In the outstanding final Office Action, the Examiner rejected claims 1-22 under 35 U.S.C. §103(a) as being unpatentable over Garg et al., "Frame-Dependent Multi-Stream Reliability Indicators for Audio-Visual Speech Recognition," (hereinafter "Garg") in view of U.S. Patent Application Publication No. 2003/0177005 to Masai et al. (hereinafter "Masai").

With regard to the issue of whether claims 1-22 are unpatentable over Garg in view of Masai, Applicants assert that the combined teaching of Garg and Masai does not result in Applicants' invention as recited in the subject claims, and Applicants' invention as recited in the subject claims is not obvious in view of the combined teaching of Garg and Masai. The Examiner contends that the combination of Garg and Masai discloses all of the claim limitations recited in the subject claims. Applicants respectfully assert that such combination fails to establish a prima facie case of obviousness, see M.P.E.P. §2143 in that the cited combination fails to teach or suggest all the claim limitations.

The present invention, for example, as recited in independent claim 1, recites a method of using a computer processor to improve speech recognition performance in an audio-visual speech recognition system. At least one of audio data and visual data associated with an input spoken utterance is received. The computer processor is used to select between an acoustic-only data model and an acoustic-visual data model based on a condition associated with a visual environment. The computer processor is also used to decode at least a portion of the at least one of audio data and visual data associated with the input spoken utterance using the selected data model. Independent claims 10, 19 and 21 recite similar limitations.

Advantageously, as illustratively explained in the present specification at page 2, during periods of degraded visual conditions, the audio-visual speech recognition system is able to decode (recognize) input speech data using audio-only data, thus avoiding recognition inaccuracies that may result from performing speech recognition based on acoustic-visual data models and degraded visual data.

Furthermore, as illustratively explained in the present specification at page 2, principles of the invention may be extended to speech recognition systems in general such that model selection (switching) may take place at the frame level. Switching may occur between two or more models. By way of example, independent claim 22 recites a method for use in accordance with a speech recognition system for improving a recognition performance thereof, comprising the steps of selecting for a given frame between a first data model and at least a second data model based on a given condition, and decoding at least a portion of an input spoken utterance for the given frame using the selected data model.

Garg, as explained in its Abstract on page 24, investigates the use of local, frame-dependent reliability indicators of the audio and visual modalities, as a means of estimating stream components of multi-stream hidden Markov models (HMM) for audio-visual speech recognition system. More specifically, Garg proposes using soft weights on each of the audio and visual HMM modalities. The value of this weight is determined through a likelihood ratio test based on observations in the acoustic space only. The dispersion metric is based on speech class conditional likelihoods, in this case, speech context dependent of independent phonemes.

As admitted by the Examiner, Garg does not specifically teach that a data model is selected based on a condition associated with the environment of the speaker. The Examiner contends that the deficiencies of Garg are remedied by Masai, which discloses selection of an acoustic model for recognition according to environment information.

Applicants assert that Garg fails to disclose selecting between an acoustic-only data model and an acoustic-visual data model based on a condition associated with a visual environment, and decoding at least a portion of at least one of audio data and visual data associated with an input spoken utterance using the selected data model, as recited in independent claims 1, 10, 19 and 21. Further, Garg fails to disclose selecting for a given frame between a first data model and at least a second data model based on a given condition, and decoding at least a portion of an input spoken utterance for the given frame using the selected data model, as recited in independent claim 22.

These deficiencies of Garg are not remedied by Masai. While Masai describes selection of an acoustic model, Masai contains no disclosure relating to a selection between an acoustic-only model and

an acoustic-visual model. Further, while Masai selects a model based on environment information, the environment information is defined as a time, place, physical condition of the speaker, etc. Thus, the environment information does not relate to a general acoustic or visual environment but instead the effect a specific time or place has on acoustics, due to the fact that the selection performed is between two acoustic models. Thus, Masai fails to disclose that the selection of a model is based on a condition associated with a visual environment. Finally, Masai fails to disclose that a model is selected based on a condition associated with an environment (visual) that acts as an input to one model (acoustic-visual data model) and does not act as an input to another model (acoustic-only data model). Should the condition be unfavorable, the model without that input is selected.

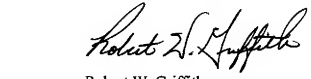
Therefore, since neither Garg nor Masai individually teach or suggest the limitations of the independent claims of the present invention as described above, the combined teaching of Garg and Masai fails to disclose these limitations, and these limitations are not obvious in view of their combined teaching. For at least these reasons, Applicants assert that independent claims 1, 10, 19, 21 and 22 are patentable over the combination of Garg and Masai.

In response to arguments previously set forth by Applicants, the Examiner contends that it is well known in the art to provide a means for selecting an optimum data model for performing recognition based on environmental conditions so as to improve recognition accuracy and performance. Applicants respectfully disagree. Masai only describes selection of an acoustic data model in accordance with surrounding acoustics, not general environmental conditions as the Examiner contends. Thus, the Examiner has failed to provide any evidence that selection between an acoustic-only data model and an acoustic-visual data model on a condition associated with a visual environment is well known in the art or obvious.

Dependent claims 2-9, 11-18 and 20 are patentable over the combination of Garg and Masai at least by virtue of their dependency from independent claims 1, 10 and 19, and also recite patentable subject matter in their own right. For example, dependent claims 2-9, 11-18 and 20 recite limitations pertaining to the model selection step/operation. However, since Garg fails to disclose a model selection step/operation, Garg is also silent regarding the details of a model selection step/operation. Further, claims 2, 11 and 20 recite storing the acoustic-only data model and the acoustic-visual data model in memory such

that model selection is made by shifting one or more pointers to one of more memory locations where the selected model is located. Despite the assertion to the contrary in the Office Action, Garg is completely silent as to any pointer shifting operation. Accordingly, withdrawal of the rejections of claims 1-22 under §103(a) is respectfully requested. In view of the above, Applicants believe that claims 1-22 are in condition for allowance, and respectfully request withdrawal of the §103(a) rejection.

Respectfully submitted,



Date: December 29, 2006

Robert W. Griffith
Attorney for Applicant(s)
Reg. No. 48,956
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-4547